IN THE CLAIMS:

1. (Currently amended) A method for tracking multiple objects in a video sequence comprising:

selecting an initial configuration comprising a plurality of objects, where the initial configuration represents an initial relationship among the objects in said plurality of objects;

predicting a current configuration, where the current configuration represents a current relationship among the objects in said plurality of objects; and

computing a likelihood for the current configuration.

2. (Original) The method of claim 1 wherein said predicting step comprises performing an object level prediction.

3. (Original) The method of claim 1 wherein said predicting step comprises performing a configuration level prediction.

4. (Original) The method of claim 3 wherein said configuration level prediction handles object addition and deletion from a current configuration.

5. (Original) The method of claim 1 wherein the predicting step comprises:

determining a percentage of the objects that are covered by the current configuration;

determining a number of current configurations that correspond to the objects; and

maximizing said percentage and minimizing said number to identify an optimal current configuration.

6. (Previously presented) The method of claim 5 wherein said percentage determining step comprises solving:

$$\gamma = \frac{\left| A \cap (\bigcup_{i=1}^{m} B_i) + b \right|}{|A| + b}$$

7. (Previously presented) The method of claim 5 wherein said percentage determining step comprises solving:

$$\xi = \frac{\left| A \cap (\bigcup_{i=1}^{m} B_i) + c \right|}{(\left| \bigcup_{i=1}^{m} B_i \right| + a)}$$

8. (Original) The method of claim 1 wherein multiple objects in a video sequence are represented by said configuration comprising a plurality of modeled objects.

9. (Original) A method of producing probability distributions of states for multiple objects in a video sequence comprising:

performing hierarchical sampling of at least one frame of video in said video sequence, wherein said sampling is performed in an object configuration and individual object states, said object configuration representing a relationship among a plurality of objects within a scene; and

repeating said sampling for each frame of video in said video sequence to track objects within the video sequence.

10. (Canceled)

11. (Currently amended) A computer readable medium containing a program that, when executed by a processor, causes an image processing system to perform a method comprising:

selecting an initial configuration comprising a plurality of objects, where the initial configuration represents an initial relationship among the objects in said plurality of objects;

predicting a current configuration, where the current configuration represents a current relationship among the objects in said plurality of objects; and

computing a likelihood for the current configuration.

12. (Original) The method of claim 11 wherein said predicting step comprises performing an object level prediction.

13. (Original) The method of claim 11 wherein said predicting step comprises performing a configuration level prediction.

14. (Original) The method of claim 13 wherein said configuration level prediction handles object addition and deletion from a current configuration.

15. (Original) The method of claim 11 wherein the predicting step comprises:

determining a percentage of the objects that are covered by the current configuration;

determining a number of current configurations that correspond to the objects;

maximizing said percentage and minimizing said number to identify an optimal current configuration.

16. (Previously presented) The method of claim 15 wherein said percentage determining step comprises solving:

$$\gamma = \frac{\left| A \cap (\bigcup_{i=1}^{m} B_i) + b \right|}{|A| + b}$$

17. (Previously presented) The method of claim 15 wherein said percentage determining step comprises solving:

$$\xi = \frac{\left| A \cap (\bigcup_{i=1}^{m} B_i) + c \right|}{(\left| \bigcup_{i=1}^{m} B_i \right| + a)}$$

18. (Original) The method of claim 11 wherein multiple objects in a video sequence are represented by said configuration comprising a plurality of modeled objects.